

NTUTH-99-033

June 1999

Fermion determinant and chiral anomaly on a finite lattice

Ting-Wai Chiu

Department of Physics, National Taiwan University

Taipei, Taiwan 106, Republic of China.

E-mail : twchiu@phys.ntu.edu.tw

Abstract

The fermion determinant and the chiral anomaly of lattice Dirac operator D on a finite lattice are investigated. The condition for D to reproduce correct chiral anomaly at each site of a finite lattice for smooth background gauge fields is that D possesses exact zero modes satisfying the Atiyah-Singer index theorem. This is also the necessary condition for D to have correct fermion determinant (ratio) which plays the important role of incorporating dynamical fermions in the functional integral. We outline a scheme for dynamical fermion simulation of lattice QCD with topologically proper D .

PACS numbers: 11.15.Ha, 11.30.Fs, 11.30.Rd

1 Introduction

The fermion determinant, $\det(D)$, of a Dirac operator D , is proportional to the exponentiation of the one-loop effective action which is the summation of any number of external sources interacting with one internal fermion loop. It is one of the most crucial quantities to be examined in any lattice fermion formulations. In numerical simulations of lattice QCD, its ratio between two successive gauge configurations determines whether the effects of dynamical quarks are properly incorporated, which are believed to be crucial to bring the *ab initio* predictions of lattice QCD to agree with the experimental data. Another important quantity in QCD is the chiral anomaly which breaks the global chiral symmetry of massless QCD through the internal quark loop. In the case of the axial current coupling to two external photons, the axial anomaly [1] provides a proper account of the decay rate of $\pi^0 \rightarrow \gamma\gamma$. For the flavor singlet axial current coupling to two gluons, the presence of chiral anomaly but no associated Goldstone boson posed the $U_A(1)$ problem which was resolved by 't Hooft [2], after taking into account of the topologically nontrivial gauge configurations, the instantons. This may also provide an explanation why the η (η') particle is much heavier than the π 's (K 's). Lattice QCD should be designed to provide nonperturbative and quantitative answers to all these and related problems. However, these goals could not be attained if the lattice Dirac fermion operator does not reproduce correct chiral anomaly or the fermion determinant (ratio) on a finite lattice *without fine-tuning any parameters*. Furthermore, lattice QCD should realize the spontaneous breaking of chiral symmetry in massless QCD, which gives massless Goldstone bosons and thus leads to a correct and detailed description of the properties of the strong interactions at low energy [3].

A basic requirement for lattice Dirac operator D is that, on a finite lattice with any prescribed¹ smooth background gauge field which has a well-defined topological charge, D possesses *exact* zero modes satisfying the Atiyah-Singer index theorem (i.e., D is *topologically proper*) without fine tuning any parameters. Here we would like to emphasize the importance of having exact zero modes on a finite lattice without fine tuning any parameters. If the zero modes can only be obtained by fine tuning parameters, then it is impossible to tune the parameters for every gauge configuration generated during a Monte Carlo simulation.

For any D satisfying this requirement, the sum of the anomaly function² over all sites must be correct, and is equal to two times of the topological charge of the background gauge field. If it turns out that the anomaly function of D at some of the sites do not agree with the Chern-Pontryagin density, we can

¹An example of prescribed background gauge field is given in the Appendix.

²Here we have adopted the terminology used in Chapter 22 of ref. [3].

perform the following *topologically invariant* transformation

$$\mathcal{T}(R) : \quad D' = \mathcal{T}(R)[D] \equiv D(\mathbb{1} + RD)^{-1} \quad (1)$$

with some operator R such that D' is local, then the anomaly function of D' would be in good agreement with the Chern-Pontryagin density at each site³. It suffices to choose R to be local in the position space and trivial in the Dirac space. The terminology, *topologically invariant*, refers to the property that *the zero modes and the index of D are invariant under the transformation (1)*. Note that it is not necessary to perform any fine tunings of R since any local D' would give the correct chiral anomaly. Evidently, the set of transformations, $\{\mathcal{T}(R)\}$, with the group multiplication defined by

$$(\mathcal{T}(R_1) \circ \mathcal{T}(R_2))[D] = \mathcal{T}(R_1)[\mathcal{T}(R_2)[D]], \quad (2)$$

form an abelian group with the group parameter space $\{R\}$ consisting of operators local in the position space and trivial in the Dirac space, due to the following basic property,

$$\mathcal{T}(R_1) \circ \mathcal{T}(R_2) = \mathcal{T}(R_1 + R_2). \quad (3)$$

In the topologically trivial sector, D^{-1} exists and (1) becomes

$$D'^{-1} = D^{-1} + R. \quad (4)$$

Then there exists a unique

$$R_c = -\frac{1}{2}\{D^{-1}, \gamma_5\}\gamma_5 \quad (5)$$

such that

$$D_c^{-1} = D^{-1} + R_c \quad (6)$$

is chirally symmetric, i.e.,

$$\gamma_5 D_c^{-1} + D_c^{-1} \gamma_5 = 0. \quad (7)$$

From the physical point of view, we must require the existence of a chirally symmetric $D_c = \mathcal{T}(R_c)[D]$ such that in the free fermion limit $D_c(p) \rightarrow i\gamma_\mu p_\mu$ as $p \rightarrow 0$, otherwise, D is irrelevant to massless QCD. In general, we assume that there exists

$$D_c = D(\mathbb{1} + R_c D)^{-1} \quad (8)$$

³Here we have assumed that the size of the finite lattice is large enough such that the finite size effects can be neglected, i.e., the size of the lattice is much larger than the localization length of D' .

such that

$$D_c \gamma_5 + \gamma_5 D_c = 0. \quad (9)$$

If $D(D_c)$ is topologically proper, then $D(D_c)$ must be free of species doubling, thus D_c is non-local, according to the Nielsen-Ninomiya no-go theorem [5]. Furthermore, D_c has singularities in topologically nontrivial background gauge fields [6]. Substituting (8) into (9), one obtains

$$D\gamma_5 + \gamma_5 D = -2DR_c\gamma_5 D. \quad (10)$$

This is the rejuvenated Ginsparg-Wilson relation [7]. In general, for *any* lattice Dirac operator D , if the chirally symmetric $D_c = \mathcal{T}(R_c)[D]$ exists, then D must satisfy the Ginsparg-Wilson relation. From this viewpoint, the Ginsparg-Wilson relation really does *not* specify the important attribute of a Dirac operator, namely its *topological characteristics* [6]. Only for D is *topologically proper* (e.g., the overlap-Dirac operator [8]), D (or its transform $D' = D(\mathbb{I} + RD)^{-1}$) can be used for lattice QCD, then the attractive features pointed out in refs. [9, 10] can be realized on a finite lattice. If D is topologically proper but nonlocal, the transformation (1) can be used to obtain a local D' such that the anomaly function at each site of a finite lattice could agree with the corresponding Chern-Pontryagin density for smooth background gauge fields. Moreover, the locality of D' also implies that its fermion determinant ratio (with the zero modes omitted) would agree with the corresponding value in continuum.

2 The Anomaly Function

In general, we consider lattice Dirac operator D which breaks the continuum chiral symmetry according to

$$D\gamma_5 + \gamma_5 D = B \quad (11)$$

where B is a generic operator which is usually taken to be irrelevant (i.e., vanish in the limit $a \rightarrow 0$). However, we can construct the chirally symmetric action

$$\mathcal{A}_s = \bar{\psi} D_s \psi \quad (12)$$

using the chirally symmetric part of D ,

$$D_s = \frac{1}{2}(D - \gamma^5 D \gamma^5). \quad (13)$$

Here the Dirac, flavor and internal symmetry indices are all suppressed. Then \mathcal{A}_s has the usual chiral symmetry, and the divergence of the associated Noether current⁴ is

$$\partial^\mu J_\mu^5(x) = \bar{\psi}_x \gamma^5 (D\psi)_x + (\bar{\psi}D)_x \gamma^5 \psi_x - \frac{1}{2}(\bar{\psi}B)_x \psi_x - \frac{1}{2}\bar{\psi}_x (B\psi)_x \quad (14)$$

which satisfies the conservation law,

$$\sum_x \partial^\mu J_\mu^5(x) = 0. \quad (15)$$

As usual, the lattice is taken to be finite with periodic boundary conditions, and $\partial_\mu J_\mu^5(x)$ is defined by the backward difference of the axial current

$$\partial_\mu J_\mu^5(x) = \sum_\mu [J_\mu^5(x) - J_\mu^5(x - \hat{\mu})] \quad (16)$$

such that it is parity even under the parity transformation, and the conservation law Eq. (15) is also satisfied.

If D does not possess exact zero modes in the background gauge field, then the fermionic average of $\partial^\mu J_\mu^5(x)$ can be evaluated as

$$\langle \partial^\mu J_\mu^5(x) \rangle = \frac{1}{Z} \int [d\psi][d\bar{\psi}] \partial^\mu J_\mu^5(x) \exp(-\bar{\psi}D\psi) \quad (17)$$

$$Z = \int [d\psi][d\bar{\psi}] \exp(-\bar{\psi}D\psi) \quad (18)$$

Using (14), we obtain

$$\langle \partial^\mu J_\mu^5(x) \rangle = \frac{1}{2} \text{tr} [(BD^{-1})(x, x) + (D^{-1}B)(x, x)] \equiv \mathcal{A}_n(x) \quad (19)$$

where the trace runs over the Dirac, flavor and internal symmetry indices. The RHS of Eq. (19) is identified to be the anomaly function of D . If the chirally symmetric limit of (1), i.e., $D_c = \mathcal{T}(R_c)[D]$ exists, then D satisfies the Ginsparg-Wilson relation (10) and the anomaly function becomes

$$\mathcal{A}_n(x) = -\text{tr} [\gamma_5(R_c D)(x, x) + \gamma_5(D R_c)(x, x)] \quad (20)$$

However, for any D which does not possess exact zero modes in the background gauge field, the sum of the anomaly function (19) over all sites must vanish due to the conservation law, Eq. (15),

$$\sum_x \mathcal{A}_n(x) = 0. \quad (21)$$

This implies that if $\mathcal{A}_n(x)$ is not zero identically for all x , then it must fluctuate from positive to negative values with respect to x . The latter case is exactly

⁴Here we assume that it is the flavor singlet of the flavor symmetry group.

what happens to the anomaly function of the standard Wilson-Dirac fermion operator.

On the other hand, if D possesses exact zero modes in topologically non-trivial background gauge fields, then D^{-1} is not well defined. In this case, one needs to introduce an infinitesimal mass m and evaluate (17) with D replaced by $\hat{D} = D + mf[D]$ (where $f[D]$ is any functional of D , which has eigenvalue one for the exact zero modes of D), and finally take the limit ($m \rightarrow 0$), i.e.,

$$\langle \partial^\mu J_\mu^5(x) \rangle = \lim_{m \rightarrow 0} \frac{1}{Z} \int [d\psi][d\bar{\psi}] \partial^\mu J_\mu^5(x) \exp(-\bar{\psi} \hat{D} \psi) \quad (22)$$

$$Z = \int [d\psi][d\bar{\psi}] \exp(-\bar{\psi} \hat{D} \psi) \quad (23)$$

Inserting (14) into (22), we obtain

$$\begin{aligned} \langle \partial^\mu J_\mu^5(x) \rangle &= \lim_{m \rightarrow 0} \frac{1}{2} \text{tr} \left[(B \hat{D}^{-1})(x, x) + (\hat{D}^{-1} B)(x, x) \right] \\ &\quad + 2 \sum_{s=1}^{N_+} [\phi_s^+(x)]^\dagger \phi_s^+(x) - 2 \sum_{t=1}^{N_-} [\phi_t^-(x)]^\dagger \phi_t^-(x) \end{aligned} \quad (24)$$

where ϕ_s^+ and ϕ_t^- are normalized eigenfunctions of D with eigenvalues $\lambda_s = \lambda_t = 0$ and chiralities $+1$ and -1 respectively. The first term on the RHS of Eq. (24) is identified to be the anomaly function of D ,

$$\mathcal{A}_n(x) = \lim_{m \rightarrow 0} \frac{1}{2} \text{tr} \left[(B \hat{D}^{-1})(x, x) + (\hat{D}^{-1} B)(x, x) \right] \quad (25)$$

Then summing Eq. (24) over all sites and using Eq. (15), we obtain

$$N_- - N_+ = \frac{1}{2} \sum_x \mathcal{A}_n(x) \quad (26)$$

This is the index theorem for *any* lattice Dirac operator on a finite lattice, in *any* background gauge field. In particular, when D does not possess any exact zero modes, Eq. (26) reduces to Eq. (21). Equation (26) implies that the sum of the anomaly function over all sites must be a well defined even integer for any background gauge field, however, it does *not* necessarily imply the Atiyah-Singer index theorem for smooth background gauge fields. This can be seen as follows. Since D does not have exact zero modes in a trivial gauge background, the index of D must be proportional to the topological charge Q of the smooth background gauge field. Now, if Q is an integer, then the proportional constant must be an integer, otherwise their product in general cannot be an integer. Denoting this integer multiplier by $c[D]$, we have

$$N_- - N_+ = c[D] Q n_f \quad (27)$$

where n_f is number of fermion flavors. In particular, for $n_f = Q = 1$, $c[D] = N_- - N_+$. Here we have assumed that $c[D]$ is constant for smooth background gauge fields. This is a reasonable assumption since $c[D]$ is an intrinsic characteristics of D . However, when the gauge field becomes rough, we expect that Eq. (27) would break down. If one insists that Eq. (27) holds even for *rough* gauge configurations, then $c[D]$ cannot be an integer constant due to the highly nonlinear effects of the gauge field. It is easy to deduce that, in general, $c[D]$ is a *rational number* functional of D , which in turn depends on the gauge configuration, but it becomes an integer constant only for smooth gauge configurations. The *topological characteristics* of D , $c[D]$, was first discussed in ref. [6], and was investigated in ref. [11] for the Neuberger-Dirac operator. Equation (27) constitutes the index theorem for *any* lattice Dirac operator on a finite lattice, in *any* background gauge field with integer topological charge. We can classify D according to its response in a smooth nontrivial background gauge field. If D does not possess any zero modes, then $c[D] = 0$, D is called *topologically trivial*. If $c[D] = 1$, (27) becomes Atiyah-Singer index theorem, then D is called *topologically proper*. If $c[D]$ is not equal to zero or one, then D is called *topologically improper*.

Now we try to obtain a general expression for the anomaly function satisfying (26) and (27). Consider the gauge configuration with constant field tensors, e.g., $F_{12}^a(x) = \frac{2\pi q_1}{L_1 L_2}$, $F_{34}^a(x) = \frac{2\pi q_2}{L_3 L_4}$ and other F 's are zero, where q_1 and q_2 are integers. Then the topological charge of this configuration is

$$Q = \frac{1}{32\pi^2} \sum_x \epsilon_{\mu\nu\lambda\sigma} F_{\mu\nu}^a(x) F_{\lambda\sigma}^b(x + \hat{\mu} + \hat{\nu}) \text{tr}\{t_a t_b\} = n \ q_1 \ q_2 \quad (28)$$

which is an integer. Here t_a and t_b are generators of the internal symmetry group with the normalization $\text{tr}\{t_a t_b\} = n \ \delta_{ab}$. If D is local, then it is reasonable to assume that $\mathcal{A}_n(x)$ is constant for all x . From Eqs. (26)-(28), we can deduce that

$$\mathcal{A}_n(x) = \frac{n_f}{16\pi^2} c[D] \epsilon_{\mu\nu\lambda\sigma} F_{\mu\nu}^a(x) F_{\lambda\sigma}^b(x + \hat{\mu} + \hat{\nu}) \text{tr}\{t^a t^b\} \quad (29)$$

However, this implies that any local D would yield the correct anomaly function on a finite lattice, up to the integer factor $c[D]$. In reality, this cannot be true in general. The crux of the problem is due to the fact that when D is topological trivial, i.e. $c[D] = 0$ in the index theorem (27), the anomaly function $\mathcal{A}_n(x)$ does not necessarily vanish identically at each site, but it can fluctuate from positive to negative values provided that its sum over all sites is zero. This is the complementary solution to the homogeneous equation $\sum_x \mathcal{A}_n(x) = 0$. Using Eq. (19), we can write the general form of the complementary solution as

$$\mathcal{A}^{(h)}(x) = \sum_{\mu} [G_{\mu}(x) - G_{\mu}(x - \hat{\mu})] \equiv \partial_{\mu} G_{\mu}(x) \quad (30)$$

where $G_\mu(x)$ is any arbitrary function. Then the general solution of the anomaly function is the sum of the complementary solution (30) and the particular solution (29),

$$\begin{aligned}\mathcal{A}_n(x) &= \frac{n_f}{16\pi^2} c[D] \epsilon_{\mu\nu\lambda\sigma} F_{\mu\nu}^a(x) F_{\lambda\sigma}^b(x + \hat{\mu} + \hat{\nu}) \text{tr}\{t^a t^b\} \\ &\quad + \sum_{\mu} [G_\mu(x) - G_\mu(x - \hat{\mu})]\end{aligned}\tag{31}$$

In general, $G_\mu(x)$ depends on the background gauge field as well as the lattice Dirac operator D . If $\mathcal{A}^{(h)}(x)$ does not vanish in the presence of a constant gauge background, then $\mathcal{A}_n(x)$ must be different from the continuum chiral anomaly, hence this D should be abandoned. Only when $\mathcal{A}^{(h)}(x)$ vanishes identically for any smooth gauge background, D can reproduce the correct chiral anomaly for $c[D] = 1$ (e.g., the Neuberger-Dirac operator). On the other hand, the pathologies of the standard Wilson-Dirac operator are : $c[D] = 0$ and $\mathcal{A}^{(h)}(x) \neq 0$.

Now we can introduce local fluctuations [e.g., the sinusoidal terms in Eqs. (63)-(66)] to the background gauge field and/or increase its topological charge Q . Then $c[D]$ in Eqs. (31) and (27) may remain constant provided that the local fluctuations of the gauge fields are not too violent and/or Q is not too large. When the gauge background becomes so rough that D is non-local, then Eq. (31) is no longer valid, but the index theorem, Eq. (27) may still hold with the same $c[D]$. If we keep on increasing the roughness of the gauge background, then D will undergo a topological phase transition, and $c[D]$ in Eq. (27) will become another integer or even a fraction. Explicit examples of topological phase transition will be demonstrated in Section 4. In general, Eq. (31) constitutes the anomaly function for *any* lattice Dirac operator D which is *local* on a finite lattice.

If the chirally symmetric limit of D , i.e., $D_c = \mathcal{T}(R_c)[D]$, exists, then D satisfies the Ginsparg-Wilson relation (10), and its anomaly function (25) becomes exactly in the same form of Eq. (20). From Eqs. (31) and (20), we obtain

$$\mathcal{A}_n(x) = -\text{tr} [\gamma_5(R_c D)(x, x) + \gamma_5(D R_c)(x, x)]\tag{32}$$

$$\begin{aligned}&= \frac{n_f}{16\pi^2} c[D] \epsilon_{\mu\nu\lambda\sigma} F_{\mu\nu}^a(x) F_{\lambda\sigma}^b(x + \hat{\mu} + \hat{\nu}) \text{tr}\{t^a t^b\} \\ &\quad + \sum_{\mu} [G_\mu(x) - G_\mu(x - \hat{\mu})]\end{aligned}\tag{33}$$

which agrees with the result derived in ref. [6, 12]. For $U(1)$ lattice gauge theory with "admissible" background gauge configurations, then Eq. (33) agrees with the result derived by Lüscher [13]. For the Overlap-Dirac operator, in the weak coupling limit or the classical continuum limit, $\partial_\mu G_\mu(x) = 0$, then Eq. (33) agrees with the results derived by Kikukawa and Yamada [14], Fujikawa [15], Adams [16] and Suzuki [17], respectively.

Inserting Eq. (25) [Eq. (20)] into Eq. (26), one obtains

$$N_- - N_+ = -\frac{1}{2} \sum_x \text{tr} [\gamma_5(R_c D)(x, x) + \gamma_5(D R_c)(x, x)] \quad (34)$$

which agrees with the index theorem for Ginsparg-Wilson fermion, derived in ref. [18, 19]. Note that (34) does *not* necessarily imply the Atiyah-Singer index theorem for smooth background gauge fields with integer topological charge.

If the lattice Dirac operator D does not have exact zero modes in the presence of topologically non-trivial background gauge fields, however, it manages to possess exact zero modes in a certain limit, say, for example, the Wilson-Dirac fermion operator [4] in the classical continuum limit (i.e., zero lattice spacing with infinite number of sites in each direction). Then on any finite lattice, this lattice Dirac operator cannot reproduce the correct anomaly function for topologically non-trivial background gauge fields, since the sum of the anomaly function (19) over the entire lattice must be zero due to the conservation law (15),

$$\frac{1}{2} \sum_x \text{tr} [(B D^{-1})(x, x) + (D^{-1} B)(x, x)] = \sum_x \langle \partial^\mu J_\mu^5(x) \rangle = 0 \quad (35)$$

while the sum of Chern-Pontryagin density over all sites is proportional to the topological charge,

$$\sum_x \frac{1}{16\pi^2} \epsilon_{\mu\nu\lambda\sigma} F_{\mu\nu}^a(x) F_{\lambda\sigma}^b(x + \hat{\mu} + \hat{\nu}) \text{tr}\{t^a t^b\} = 2 Q \quad (36)$$

which is nonzero for topologically nontrivial background gauge fields. As we increase the number of sites or decrease the lattice spacing, the anomaly function does not approach the Chern-Pontryagin density since it must satisfy the constraint Eq. (35). It is evident that *the limit where D might possess exact zero modes and correct anomaly function cannot be obtained by smoothly extrapolating measurements at finite lattice spacings towards $a = 0$ since the total anomaly (35) must undergo a discontinuous jump from zero to a nonzero integer at the point $a = 0$, if the background gauge field is topologically nontrivial.* Only for the trivial background gauge field, D might have the chance to reproduce the correct index by extrapolations towards the limit $a = 0$. If one attempts to bypass this difficulty by adding a mass term to D with the hope of obtaining the correct anomaly by tuning the mass and/or other parameters, then one must encounter the unsurmountable problem of fine-tuning, since it is almost impossible to tune the parameters to some fixed values such that the Chern-Pontryagin density is reproduced at each site, and for all different gauge configurations. Therefore the chiral limit of these lattice Dirac operators (e.g., the standard Wilson-Dirac fermion operator) not only is difficult to be realized in practice, but also in principle a goal impossibly to be attained. On

the other hand, if the lattice Dirac operator (e.g., the overlap-Dirac operator [8]) can reproduce exact zero modes satisfying the Atiyah-Singer index theorem on a finite lattice, then the total anomaly is bound to be correct, thus the anomaly function could agree with the Chern-Pontryagin density even at finite lattice spacings.

3 Fermion Determinant

Now let us turn to the fermion determinant which plays the important role of incorporating dynamical fermions in the functional integral. Formally, the fermion determinant can be represented as the product of all eigenvalues of D . In continuum, the fermion determinant must be regularized, and only the ratio of fermion determinants can be well defined and independent of regularization. In the topologically trivial sector, it is usually normalized to one in the free fermion limit, i.e., $\det[D(A)] \leftarrow \det[D(A)]/\det[D(0)]$. However, in the non-trivial sectors, D has exact zero modes, then the fermion determinant ratio is well-defined only for gauge configurations within the same topological sector or with the zero modes omitted.

On a lattice, if D is topologically proper, then the fermion determinant in the functional integral also serves another important role to "*decompose*" configurations of link variables into distinct topological sectors. This can be seen as follows. On a lattice, any two gauge configurations are smoothly connected. In Monte Carlo simulations, the transition probability between any two gauge configurations in the quenched approximation is proportional to the exponentiation of the difference of their actions, which is finite. This is in contrast to the situation in the continuum, where the gauge configurations are decomposed into disconnected topological sectors, and the functional integral is performed in each topological sector individually. However, if D is topologically proper and its fermion determinant is incorporated in the functional integral, then the transition probability between any two gauge configurations in two different topological sectors is either *zero* or *infinite*, in the limit $m \rightarrow 0$ (Here m is an infinitesimal mass introduced to D). So, if one cutoffs all transitions having a transition probability less than a certain lower bound, say, $p_1(m)$, or larger than an upper bound, say, $p_2(m)$, then only gauge configurations within the same topological sector can be accepted in the Monte Carlo updatings, thus the functional integral can be evaluated exactly in the same manner as in the continuum with disconnected topological sectors. Therefore Monte Carlo simulations and measurements of observables are performed separately for each relevant topological sector, and the final result is averaged over all relevant topological sectors with weight $e^{i\nu\theta}$, where θ denotes the vacuum angle and ν the winding number. From this viewpoint, the deficiency of link variables being smoothly deformable to the identity element can be circumvented by

incorporating dynamical fermions in the functional integral, provided that D is topologically proper.

To be specific, consider lattice QCD with n_f flavors of "massless" quarks, whereas the lattice Dirac operator D is topologically proper. Then the expectation value of a general observable $\mathcal{O}(U, \psi, \bar{\psi})$ of the field variables can be written as

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int [dU][d\psi][d\bar{\psi}] \mathcal{O}(U, \psi, \bar{\psi}) \exp(-\mathcal{A}_g - \bar{\psi} D(U) \psi) \quad (37)$$

where \mathcal{A}_g is Wilson's pure gauge action, and Z is the partition function

$$Z = \int [dU] \det[D(U)] \exp(-\mathcal{A}_g) \quad (38)$$

Here the functional integral of gauge fields must include the contributions of instantons. In four dimensions, for any gauge group G which is a simple Lie group (e.g., $G = SU(3)$), the instantons are characterized by the elements of the homotopy group $\pi_3(G) = \pi_3(SU(2)) = \mathbb{Z}$, i.e., the winding number,

$$\nu = \frac{1}{64\pi^2} \sum_x \epsilon_{\mu\nu\lambda\sigma} F_{\mu\nu}^a(x) F_{\lambda\sigma}^a(x + \hat{\mu} + \hat{\nu}) \quad (39)$$

which satisfies

$$n_f \nu = N_- - N_+ \quad (40)$$

where $N_+(N_-)$ denotes the number of zero modes of D of positive (negative) chirality. Suppose that the weight factor for each topological sector is $w(\nu)$, then Eqs. (37) and (38) are replaced by

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \sum_{\nu} w(\nu) \int [dU][d\psi][d\bar{\psi}] \mathcal{O}(U, \psi, \bar{\psi}) \exp(-\mathcal{A}_g - \bar{\psi} D(U) \psi) \quad (41)$$

$$Z = \sum_{\nu} w(\nu) \int [dU] \det[D(U)] \exp(-\mathcal{A}_g) . \quad (42)$$

If we require $\langle \mathcal{O} \rangle$ to satisfy the cluster decomposition principle [3], then it is straightforward to deduce that

$$w(\nu) = \exp(i\nu\theta) \quad (43)$$

where θ is a free parameter, the so called vacuum angle [20]. Due to the presence of fermion determinant, the partition function (42) only receives the contribution from the trivial sector,

$$Z = \int [dU]^{(0)} \det[D(U)] \exp(-\mathcal{A}_g) \quad (44)$$

where the superscript (0) stands for the trivial sector. Without loss of generality, the observable \mathcal{O} can be written as

$$\mathcal{O}(U, \psi, \bar{\psi}) = \mathcal{O}(U) \psi_{x_1} \bar{\psi}_{y_1} \cdots \psi_{x_n} \bar{\psi}_{y_n} . \quad (45)$$

Then, for the trivial sector, (41) can be evaluated as

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \sum_{\nu} e^{i\nu\theta} \int [dU] \exp(-\mathcal{A}_g) \mathcal{O}(U) \det(D) \sum_{z_1, \dots, z_n} \epsilon_{x_1 \dots x_n}^{z_1 \dots z_n} D_{z_1 y_1}^{-1} \cdots D_{z_n y_n}^{-1} \quad (46)$$

where $\epsilon_{x_1 \dots x_n}^{z_1 \dots z_n}$ is the completely antisymmetric tensor which is equal to $+1(-1)$ if $\{z_1 \cdots z_n\}$ is even (odd) permutation of $\{x_1 \cdots x_n\}$, and is zero otherwise. For topologically non-trivial sectors, (41) can be evaluated by expanding the fermion fields in terms of the complete set of eigenfunctions of D with Grassman coefficients, then the functional integral over the fermion fields can be expressed in terms of that over the Grassman coefficients. It is evident that for \mathcal{O} given in (45), non-trivial sectors with winding number $|\nu| > n$ do not contribute to (41). In general, (41) can be written as

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \sum_{\nu} \exp(i\nu\theta) \int [dU] \exp(-\mathcal{A}_g) \det_1[D(U)] \mathcal{O}(U) \mathcal{F}(U) \quad (47)$$

where $\det_1[D(U)]$ denotes the fermion determinant with the zero modes omitted, and $\mathcal{F}(U)$ the remnants of the observable involving fermion fields after incorporating the zero modes from the fermion determinant, which in general can be expressed in terms of the eigenfunctions of the zero modes of D . It is evident that the relevant topological sectors which have nonzero contributions to (47) must come in $\pm\nu$ pairs ($\nu \neq 0$). It is reasonable to expect that the functional integrals for $\pm\nu$ sectors are equal, especially for observables in vector gauge theories such as QCD.

Now consider Monte Carlo simulations of lattice QCD with dynamical quarks. Suppose the initial gauge configuration is $\{U\}$ in the topological sector of winding number ν . A trial configuration $\{U'\}$ with winding number ν' is generated. We intend to reject $\{U'\}$ if $\nu' \neq \nu$. Then, for $\nu' = \nu$, $\{U'\}$ is accepted according to the Metropolis algorithm. That is, we compute the ratio

$$P(\{U\} \rightarrow \{U'\}) \equiv \frac{\det[D(U')]}{\det[D(U)]} \exp(-\mathcal{A}_g(U') + \mathcal{A}_g(U)) \quad (48)$$

and compare P with a random number x drawn from the uniform distribution in $(0, 1)$. If $x < P$, then $\{U'\}$ is accepted, otherwise $\{U'\}$ is rejected. Now our problem is how to reject $\{U'\}$ with $\nu' \neq \nu$, before the Metropolis sampling.

In general, for any two configurations $\{U\}$ and $\{U'\}$, (48) can be written as

$$P(\{U\} \rightarrow \{U'\}) = 0^{n_f(|\nu'| - |\nu|)} \frac{\det_1[D(U')]}{\det_1[D(U)]} \exp(-\mathcal{A}_g(U') + \mathcal{A}_g(U)) \quad (49)$$

where \det_1 denotes the fermion determinant with zero modes omitted⁵. It is

⁵Here we have assumed that D is normal and thus $\det_1(D)$ is positive definite.

evident that P is zero if $|\nu'| > |\nu|$, but ∞ if $|\nu'| < |\nu|$. So, we can easily reject $\{U'\}$ in these two cases. Next we determine whether $\nu' = \nu$ or $\nu' = -\nu$. This can be done in the following. In reality, the quarks may not be exactly massless. However, if there exists a very light quark (e.g. u or d quark) of almost zero mass which enters the action in the form $m\bar{\psi}_R\psi_L + m^*\bar{\psi}_L\psi_R$ (m^* is complex conjugate of m), then the factor $0^{n_f(|\nu'| - |\nu|)}$ in Eq. (49) is replaced by

$$R(m) \equiv \frac{m^{\nu'}\theta(\nu') + m^{*- \nu'}\theta(-\nu')}{m^\nu\theta(\nu) + m^{*- \nu}\theta(-\nu)} \quad (50)$$

where $\theta(x)$ is the step function. In order to reject the configurations $\{U'\}$ with winding number $\nu' \neq \pm\nu$, we first set $m^* = m \sim 0$, and compute P according to its definition (48). Then we reject $\{U'\}$ if $P < p_1(m)$ or $P > p_2(m)$, where $p_1(m)$ and $p_2(m)$ are easily determined since P becomes either zero or infinite for $\nu' \neq \pm\nu$, in the limit $m \rightarrow 0$. Now we have $\{U'\}$ with $\nu' = \pm\nu$. Next we restore $m^* \neq m$ and compute P again using (48). If P is real, then $\nu' = \nu$, according to (50), so $\{U'\}$ is further sampled by Metropolis algorithm. On the other hand, if P is complex, then $\nu' = -\nu$, but the configuration $\{U'\}$ can be used for the sector of winding number $-\nu$ rather than being thrown away. As mentioned above, in general, the relevant topological sectors for any observable must come in $\pm\nu$ pairs. So, our strategy is to perform Monte Carlo simulations for both $\pm\nu$ sectors simultaneously such that optimal efficiency can be attained. After enough statistics have been accumulated for $\pm\nu$ sectors, then we can move on to another pair of relevant topological sectors by starting with a prescribed gauge configuration. Finally the expectation value $\langle \mathcal{O} \rangle$, (47), can be obtained by averaging measurements over all relevant topological sectors with weight factor $\exp(i\nu\theta)$.

On the other hand, if D does *not* possess exact zero modes, the above prescription is no longer valid, then we cannot decompose the link configurations into distinct topological sectors, in contrast with the gauge configurations in continuum. Furthermore, if D is topologically improper (or trivial), then its fermion determinant must be very different from its counterpart in continuum. Even if we introduce a mass term, the fermion determinant of these Dirac operators (e.g., the standard Wilson-Dirac fermion operator⁶) does not tend to the correct value in the chiral limit, thus the effects of dynamical fermion cannot be incorporated properly. Therefore the *necessary* condition for D (or $D' = D(\mathbb{1} + RD)^{-1}$) to have the correct fermion determinant is that D is topologically proper. Then the locality of D' would in general lead to the correct fermion determinant $\det(D')$ on a finite lattice. However, we still cannot formulate a criterion (such as the locality of a topologically proper D')

⁶We recall that the Wilson-Dirac fermion operator is topologically trivial, which does not possess exact zero modes on a finite lattice.

that can guarantee the correct fermion determinant on a finite lattice. We are sure that, if such a criterion exists, it must be more stringent than that for the chiral anomaly, since the former involves all eigenvalues of D while the latter only the zero modes.

A viable scheme to reproduce the continuum fermion determinant on the lattice is Narayanan and Neuberger's Overlap formalism [21] which was inspired by Kaplan's Domain-Wall fermion [22] and Frolov-Slavnov's generalized Pauli-Villars regularization [23] with an infinite number of Pauli-Villars fields. The lattice Dirac operator proposed by Neuberger [8] fulfils the crucial requirements of reproducing correct chiral anomaly and fermion determinant on a finite lattice, as reported in ref. [24].

4 Some Examples

In this section, we illuminate above discussions by computing the axial anomaly and the fermion determinant of a topologically proper D which is constructed for the purpose of demonstration, and compare with those of the standard Wilson-Dirac fermion operator which is topologically trivial. For simplicity, we perform these computations on two-dimensional lattices with $U(1)$ background gauge field, and compare the numerical results with the exact solution on the torus. It is straightforward to extend these numerical computations to four dimensional lattices with non-abelian background gauge fields.

First, we set up the notations for the Wilson-Dirac fermion operator D_w with Wilson parameter $r_w > 0$,

$$D_w = m_f + \frac{1}{2} [\gamma_\mu (\nabla_\mu^* + \nabla_\mu) - r_w \nabla_\mu^* \nabla_\mu] \quad (51)$$

where m_f is the fermion mass, ∇_μ and ∇_μ^* are the forward and backward difference operators defined as follows.

$$\begin{aligned} \nabla_\mu \psi(x) &= U_\mu(x) \psi(x + \hat{\mu}) - \psi(x) \\ \nabla_\mu^* \psi(x) &= \psi(x) - U_\mu^\dagger(x - \hat{\mu}) \psi(x - \hat{\mu}) \end{aligned}$$

From Eq. (19), the divergence of axial current for the Wilson-Dirac fermion operator is

$$\begin{aligned} \langle \partial^\mu J_\mu^5(x) \rangle &= 2m_f \text{tr} [\gamma^5 D_w^{-1}(x, x)] \\ &\quad - \frac{r_w}{2} \text{tr} \left\{ \gamma^5 [(\nabla_\mu^* \nabla_\mu) D_w^{-1} + D_w^{-1} (\nabla_\mu^* \nabla_\mu)](x, x) \right\} \end{aligned} \quad (52)$$

The last term in (52) is the axial anomaly which can be written as

$$\begin{aligned} \mathcal{A}_n(x) &= \frac{r_w}{2} \sum_\mu \left\{ 4 \text{tr} [D_w^{-1}(x, x) \gamma^5] - \text{tr} [D_w^{-1}(x, x + \hat{\mu}) \gamma^5 U_\mu(x)] \right. \\ &\quad - \text{tr} [D_w^{-1}(x + \hat{\mu}, x) \gamma^5 U_\mu^\dagger(x)] - \text{tr} [D_w^{-1}(x - \hat{\mu}, x) \gamma^5 U_\mu(x - \hat{\mu})] \\ &\quad \left. - \text{tr} [D_w^{-1}(x, x - \hat{\mu}) \gamma^5 U_\mu^\dagger(x - \hat{\mu})] \right\}. \end{aligned} \quad (53)$$

In the classical continuum limit ($a \rightarrow 0$), (53) was shown [25]-[28] to agree with the Chern-Pontryagin density. On the other hand, the sum of (53) over all sites on a finite lattice must be zero in the massless limit, as shown in Eq. (35), since D_w does not have exact zero modes. So, in general, the axial anomaly (53) of the Wilson-Dirac fermion operator on a finite lattice must be different from the Chern-Pontryagin density. Only in the classical continuum limit when the lattice spacing is equal to zero and the number of sites is infinite, then the exact zero modes could possibly emerge and the correct axial anomaly might be recovered. If this does happen, then we would expect that in the massless limit, the sum of (53) over all sites must undergo a discontinuous transition around the point $a = 0$, for topologically nontrivial gauge background.

Next, we consider a *topologically proper* Dirac operator which has exact zero modes satisfying the Atiyah-Singer index theorem. Suppose that one has constructed a lattice Dirac operator⁷ as the following

$$D = 5 \left(\sqrt{D_w^\dagger D_w} + D_w \right) \left(3\sqrt{D_w^\dagger D_w} - 2D_w \right)^{-1} \quad (54)$$

where D_w is the Wilson-Dirac operator with a negative mass parameter m_0 and Wilson parameter r_w ,

$$D_w = -m_0 + \frac{1}{2} [\gamma_\mu (\nabla_\mu^* + \nabla_\mu) - r_w \nabla_\mu^* \nabla_\mu] \quad (55)$$

For the moment, we fix $m_0 = 1$ and $r_w = 1$. Since D satisfies the Ginsparg-Wilson relation (10), the anomaly function of D can be easily obtained from Eq. (25). First, we investigate the properties of D [Eq. (54)] in background gauge fields with constant field tensor. As we expect, D possesses exact zero modes satisfying the Atiyah-Singer index theorem. However, the anomaly function is not constant, though the fluctuations are not very large. The deviation of the anomaly function of D from the Chern-Pontryagin density is defined as

$$\delta_D \equiv \frac{1}{N_s} \sum_x \frac{|\mathcal{A}_n(x) - \mathcal{A}_c(x)|}{|\mathcal{A}_c(x)|} \quad (56)$$

where the summation runs over all sites, and N_s is the total number of sites, and $\mathcal{A}_c(x)$ is the Chern-Pontryagin density which is equal to

$$\mathcal{A}_c(x) = \frac{1}{2\pi} \epsilon_{\mu\nu} F_{\mu\nu}(x) \quad (57)$$

on a torus. For constant field tensor with topological charge Q , it becomes a constant

$$\mathcal{A}_c(x) = \frac{2Q}{L_1 L_2} \quad (58)$$

⁷This is a generalized Neuberger-Dirac operator.

Q	$\mathcal{A}_c(x) = Q/(L_1 L_2)$	δ_D	$\delta_{D'}$	δ_{D_w}
1	1.389×10^{-2}	6.232×10^{-2}	2.687×10^{-4}	1.0
2	2.778×10^{-2}	2.434×10^{-2}	2.413×10^{-4}	1.0
3	4.167×10^{-2}	8.026×10^{-3}	2.036×10^{-4}	1.0
4	5.556×10^{-2}	3.192×10^{-3}	1.982×10^{-4}	1.0
5	6.944×10^{-2}	1.026×10^{-3}	1.363×10^{-4}	1.0
6	8.333×10^{-2}	5.149×10^{-4}	1.184×10^{-4}	1.0
7	9.722×10^{-2}	1.910×10^{-4}	5.873×10^{-5}	1.0
8	1.111×10^{-1}	1.061×10^{-4}	3.446×10^{-5}	1.0

Table 1: The chiral anomaly versus the topological charge Q . The exact solution is listed in the second column. The deviations of the chiral anomaly for D [Eq. (54)] and D' [Eq. (60)] are listed in the third and fourth columns respectively. The last column lists the deviations for the Wilson-Dirac fermion.

In Table 1, we list the deviation of the anomaly function of D , δ_D , for topological charge $Q = 1$ to $Q = 8$ on a 12×12 lattice. Although the deviation δ_D in the third column seems to be tiny in each case, it should be regarded as large when comparing with the deviation $\delta_{D'}$ of a local D' [Eq. (60)] listed in the fourth column. In fact, the fluctuations of the axial anomaly of D can be seen clearly by plotting $\mathcal{A}_n(x)$ versus each site of the lattice, as shown in Fig. 1. The horizontal line at height $1/72$ denotes the constant Chern-Pontryagin density, $\mathcal{A}_c(x)$ for $Q = 1$ on a 12×12 torus. The triangles denote the axial anomaly of D , and its fluctuations are due to the non-locality of D . The average of the axial anomaly of D , of course, must agree with the Chern-Pontryagin density, since its index is equal to Q . On the other hand, the axial anomaly [Eq. (53)] of the standard Wilson-Dirac fermion operator D_w [Eq. (51)] with $m_f = 0$, which is plotted as squares in Fig. 1, is very different from the Chern-Pontryagin density at each site. This is due to the fact that D_w does not have exact zero modes, thus the sum of its axial anomaly over all sites is zero on any finite lattice. So, the deviation of the axial anomaly of massless Wilson-Dirac operator, δ_{D_w} , is equal to one for any Q , as listed in the last column of Table 1.

The non-locality of D can be easily seen by plotting $D(x, y)$ as a function of $|x - y|$. In Fig. 1, one of the Dirac components of $|D(x, 0)|$, say, $|D_{11}(x, 0)|$ is plotted versus the distance $|x|$, on a 12×12 lattice with periodic boundary conditions, in a constant background gauge field with topological charge $Q = 1$. It is obvious that $D(x, y)$ is nonlocal.

Although D is topologically proper, the non-locality of D must yield incorrect chiral anomaly and fermion determinant. In Table 2, we compute the fermion determinant of D for different topological sectors of constant field

Q	$\det_1[D(Q)]_{exact}$	$\det_1[D(Q)]$	$\det_1[D'(Q)]$	$\det[D_w(Q)]$	$\det'[D_w(Q)]$
1	1.00000	1.00000	1.00000	1.00000	1.00000
2	4.24264	5.21339	4.22458	0.34182	3.98005
3	13.8564	23.5281	13.7284	0.14620	11.8619
4	38.1838	98.6168	37.2195	0.07092	28.8108
5	92.7342	394.114	90.3215	0.03841	61.1240
6	203.647	1535.83	198.435	0.02185	112.466
7	411.296	5759.04	391.581	0.01337	190.058
8	773.221	21275.2	726.985	0.00854	293.285

Table 2: The fermion determinant versus the topological charge Q . The exact solution for a torus of size 12×12 is listed in the second column, which is computed according to Eq. (59). The normalization constants are fixed by the first row in each case. The lattice Dirac operators D , D' and D_w are defined in Eqs. (54), (60) and (51) respectively.

tensor, and compare them with the exact solution on a torus [29],

$$\det_1[D(Q)]_{exact} = N \sqrt{\left(\frac{L_1 L_2}{2|Q|}\right)^{|Q|}} \quad (59)$$

where the zero modes have been omitted and the normalization constant is fixed by $N = \sqrt{\frac{2}{L_1 L_2}}$ such that $\det_1[D(1)]_{exact} = 1$. It is evident that the fermion determinant of D (in the third column) disagrees with the exact solution (in the second column), while those of a local D' [Eq. (60)] (in the fourth column) agree fairly well with the exact solutions. The fermion determinants for the Wilson-Dirac operator are listed in the fifth column, where all eigenvalues of D_w are taken into account. So, the fermion determinant must become smaller for larger Q due to the emergence of small real eigenvalues. If we discard these small real eigenvalues of D_w , which correspond to the exact zero modes of D , then the resulting fermion determinant of D_w is listed in the last column of Table 2, but still in disagreement with the exact solution. In general, the fermion determinant ratio of D_w between any two gauge configurations (no matter within the same topological sector, or in two different sectors) disagrees with the exact solution, except for the trivial sector with $Q = 0$.

As long as we have a topologically proper D , we can use the topologically invariant transformation (1) to obtain a local D' such that the anomaly function and the fermion determinant both agree with the exact solutions for smooth gauge configurations. For simplicity, we fix $R = r\mathbb{1}$ in (1). Note that there is a large range of r values to render D' local for constant background

gauge fields. For instance, we can pick $r = 0.5$, then (1) gives

$$D' = 10 \left(\sqrt{D_w^\dagger D_w} + D_w \right) \left(11 \sqrt{D_w^\dagger D_w} + D_w \right)^{-1} \quad (60)$$

It is easy to check that D' is local by plotting $|D'(x, y)|$ versus $|x - y|$, as shown in Fig. 2. The chiral anomaly and fermion determinant of D' agree very well with the exact solutions, as listed in Table 1 and Table 2.

Now we can introduce local fluctuations to the field tensor through the sinusoidal terms in Eqs. (72) and (73). For $Q = 1$, $h_1 = 0.1$, $h_2 = 0.2$, $A_1^{(0)} = 0.3$, $A_2^{(0)} = 0.4$ and $n_1 = n_2 = 1$ on a 12×12 lattice, the anomaly function of D' is computed and compared with the Chern-Pontryagin density on the torus, as shown in Fig. 4. It is evident that $A_n(x)$ agrees with $A_c(x)$ very well at each site.

Next we consider another lattice Dirac operator D'' which has the same form of D' (60) except $m_0 = 0.1$ in the D_w ,

$$D'' = 10 \left(\sqrt{D_w^\dagger D_w} + D_w \right) \left(11 \sqrt{D_w^\dagger D_w} + D_w \right)^{-1} \quad (61)$$

$$(62)$$

where

$$D_w = -0.1 + \frac{1}{2} [\gamma_\mu (\nabla_\mu^* + \nabla_\mu) - \nabla_\mu^* \nabla_\mu] .$$

Again we investigate the properties of D'' in a background gauge field of constant field tensor. Then we find that D'' does not have any zero modes for $Q \geq 5$, in contrast to the behaviors of D and D' having zero modes satisfying the Atiyah-Singer index theorem. In terms of the general index theorem, Eq. (27), we list the topological characteristics of D , D' and D'' in Table 3. It shows clearly that $c[D'']$ depends on the gauge configuration through D'' . Note that D is nonlocal for $Q = 1$ to $Q = 8$, but the Atiyah-Singer index theorem is still satisfied. We recall that the nonlocal D and the local D' are related by the topological invariant transformation (1) which preserves the zero modes as well as the index.

Now we introduce local fluctuations to the field tensor through the sinusoidal term in Eq. (72). Starting with constant field tensor at $Q = 4$, we gradually increase the amplitude of the sinusoidal term, $A_1^{(0)}$ with $n_2 = 1$. The results are shown in Table 4. For $A_1^{(0)} \leq 0.0802$, the index of D'' is 4, satisfying the Atiyah-Singer index theorem. However, around $A_1^{(0)} = 0.0803$, the index of D'' becomes 3, thus the index theorem (27) is violated for any integer $c[D'']$. Therefore, we deduce that, in general, if $c[D]$ is not a constant for any integer topological charge Q , we can generalize $c[D]$ to a rational number, then the index theorem (27) also holds for *any* rough gauge fields. It is remarkable that the index of D'' changes so abruptly around $A_1^{(0)} = 0.0802 \sim 0.0803$, a singular

Q	$c[D]$	$c[D']$	$c[D'']$
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	0
6	1	1	0
7	1	1	0
8	1	1	0

Table 3: The topological characteristics of D , D' and D'' versus the topological charge Q on a 12×12 lattice.

behavior characterizes the topological phase transition. It is very unimaginable for us to see odd number of zero modes ($N_- = 3$) in a background gauge field of even number of topological charges ($Q = 4$). The underlying non-perturbative mechanism is very intriguing, and might have many unexpected consequences.

We note in passing that if we substitute D' [Eq. (60)] and D'' [Eq. (61)] by the Neuberger-Dirac operator $D_h = \mathbb{1} + D_w(D_w^\dagger D_w)^{-1/2}$ with $m_0 = 1$ and $m_0 = 0.1$ respectively. The essential features of above numerical results are almost the same. In particular, the good agreement between the anomaly function and the Chern-Pontryagin density at each site, as shown in Fig. 4.

5 Summary and Discussions

In summary, we have shown that the condition for lattice Dirac operator D (or $D' = D(\mathbb{1} + RD)^{-1}$) to reproduce the continuum anomaly function at each site of a finite lattice for smooth background gauge fields is that D possesses exact zero modes satisfying the Atiyah-Singer index theorem (i.e., D is *topologically proper*). This is the basic requirement for any Dirac operator to be used for lattice QCD if one wishes to reproduce correct chiral anomaly and fermion determinant (ratio) on a finite lattice. It is evident that the standard Wilson-Dirac fermion operator does not meet this basic requirement. If D is topologically proper but nonlocal, then we can use (1) to transform D into a local $D' = D(\mathbb{1} + RD)^{-1}$ such that the anomaly function of D' agrees with the Chern-Pontryagin density in continuum. Note that there is no need for fine-tunings of R . The set of transformations $\{\mathcal{T}(R)\}$ defined in (1) form an abelian group with the group parameter space $\{R\}$ consisting of operators local in the position space and trivial in the Dirac space. It is evident that (1) is one of the simplest topologically invariant transformations which preserve the zero modes and the index of D .

Q	$A_1^{(0)}$	$N_- - N_+$	$c[D'']$
4	0.0100	4	1
4	0.0200	4	1
4	0.0400	4	1
4	0.0600	4	1
4	0.0800	4	1
4	0.0801	4	1
4	0.0802	4	1
4	0.0803	3	3/4
4	0.0804	3	3/4
4	0.0810	3	3/4
4	0.0820	3	3/4
4	0.0900	3	3/4

Table 4: The topological characteristics of D'' versus the amplitude of the local sinusoidal fluctuation on a 12×12 lattice, where $c[D'']$ is obtained in accordance with the general index theorem (27).

The chiral anomaly on the lattice can be understood from two seemingly different viewpoints. In general, if we have lattice Dirac operator D which breaks the usual chiral symmetry according to Eq. (11), then the chiral anomaly can be obtained through the chiral symmetry breaking term [Eq. (25)], while the fermion measure is invariant under the usual chiral transformation, in contrast to the continuum theory in which the non-invariance of the fermion measure under the chiral transformation leads to the emergence of chiral anomaly after regularization [30]. Another viewpoint to understand the chiral anomaly on a finite lattice is only for those D having chirally symmetric limit D_c under the transformation (1). Then D satisfies the Ginsparg-Wilson relation (10) which is the lattice chiral symmetry, thus the lattice action is invariant under the lattice chiral transformation. However, the fermion measure is not invariant under the lattice chiral transformation, thus leads to the chiral anomaly [19], reminiscent of the situation in continuum. Therefore, for Ginsparg-Wilson Dirac operator, one must obtain the same anomaly function (32) no matter which viewpoint one prefers. However, Eqs. (25) and (31) are the general formulas of the anomaly function for *any* lattice Dirac operator.

There is a vital difference between the continuum Dirac operator and the lattice ones. In continuum, the Dirac operator in a smooth non-trivial gauge background must have exact zero modes satisfying the Atiyah-Singer index theorem. On the other hand, lattice Dirac operator does not necessarily possess exact zero modes in a smooth non-trivial gauge background, not to mention satisfying the Atiyah-Singer index theorem. This naturally leads to the concept of topological characteristics, $c[D]$, which is an intrinsic attribute of D , in general, is *not* entirely due to the non-locality of D and/or the occurrence of

species doubling. Even if one starts with a lattice Dirac operator D which is local and free of species doubling in a constant background gauge field, and D also satisfies the Atiyah-Singer index theorem. However, as we increase the topological charge or the local fluctuations of the background gauge field, D may undergo a topological phase transition and its zero modes may violate the Atiyah-Singer index theorem. In fact D has become nonlocal long before it reaches the topological phase transition point. So, the non-locality of D cannot be the crucial factor for violations of the Atiyah-Singer index theorem, though it explains the breakdown of the anomaly function (29) unambiguously. If one tries to explain the violations of Atiyah-Singer index theorem in terms of species doubling, then one has great difficulties to explain the occurrence of odd number of zero modes in a gauge background of even number of topological charges, as shown in Table 4. Therefore we conclude that the topological characteristics of lattice Dirac operator D is a basic attribute of D , which in general is a rational number appearing in the general index theorem (27) for integer topological charge Q .

Given a lattice Dirac operator D , we cannot guarantee that for any gauge configuration, there exists an operator R such that the topologically invariant transformation (1) gives a *local* and *smooth* D' . The notion of locality is usually defined as follows : $|D(x, y)| \sim \exp(-M|x - y|/a)$ with $M = O(1)$, or $|D(x, y)| = 0$ for $|x - y| > d$ with $d = O(a)$. Although $|D(x, y)|$ can be fitted by $\exp(-M|x - y|/a)$ with $M = O(1)$, it may turn out that $|D(x, y)|$ is still highly fluctuating and/or anisotropic. So, it is necessary to introduce the notion of "smoothness" which measures the fluctuations of $|D(x, y)|$ with respect to $\exp(-M|x - y|/a)$. To determine the criterion for the gauge configuration such that D' is local and smooth is beyond the scope of the present paper. It is apparent that such criterion should take into account of the fluctuations between the plaquettes as well as the deviations of each plaquette from the identity. Otherwise the criterion is only useful in the weak field limit. For example, the Neuberger-Dirac operator D_h with $m_0 = r_w = 1$ and D' [Eq. (60)] are both local and smooth for gauge configurations with constant field tensors up to very large Q values (see Table 3), though each plaquette is very different from the identity. If one takes into account of the degree of freedom provided by the transformation (1), then the criterion for a smooth gauge configuration could be less stringent than without it.

A scheme for dynamical fermion simulation of lattice QCD is outlined in Section 3. Since the direct computation of fermion determinant is prohibitively expensive for lattice QCD, a practical implementation of the scheme must be worked out before we can proceed to any realistic calculations. Nevertheless, the prescription of decomposing link configurations into distinct topological sectors and performing Monte Carlo simulations for each relevant topological sector individually may open new avenues to tackle the longstanding problems in QCD.

A

In this appendix, we give an example of prescribed background gauge field which has been used in ref. [24, 11] for two dimensional and four dimensional lattices respectively. On a 4-dimensional torus ($x_\mu \in [0, L_\mu], \mu = 1, \dots, 4$), the simplest nontrivial gauge fields can be represented as

$$A_1(x) = t^a \left[\frac{2\pi h_1}{L_1} - \frac{2\pi q_1 x_2}{L_1 L_2} + A_1^{(0)} \sin \left(\frac{2\pi n_2}{L_2} x_2 \right) \right] \quad (63)$$

$$A_2(x) = t^a \left[\frac{2\pi h_2}{L_2} + A_2^{(0)} \sin \left(\frac{2\pi n_1}{L_1} x_1 \right) \right] \quad (64)$$

$$A_3(x) = t^a \left[\frac{2\pi h_3}{L_3} - \frac{2\pi q_2 x_4}{L_3 L_4} + A_3^{(0)} \sin \left(\frac{2\pi n_4}{L_4} x_4 \right) \right] \quad (65)$$

$$A_4(x) = t^a \left[\frac{2\pi h_4}{L_4} + A_4^{(0)} \sin \left(\frac{2\pi n_3}{L_3} x_3 \right) \right] \quad (66)$$

where t^a is any one of the generators of the gauge group with the normalization $\text{tr}(t^a t^b) = n \delta^{ab}$, and q_1 and q_2 are integers. The global part is characterized by the topological charge

$$Q = \frac{1}{32\pi^2} \int d^4x \epsilon_{\mu\nu\lambda\sigma} \text{tr}(F_{\mu\nu} F_{\lambda\sigma}) = n q_1 q_2 \quad (67)$$

which is an integer. The harmonic parts are parameterized by four real constants h_1, h_2, h_3 and h_4 . The local parts are chosen to be sinusoidal fluctuations with amplitudes $A_1^{(0)}, A_2^{(0)}, A_3^{(0)}$ and $A_4^{(0)}$, and frequencies $\frac{2\pi n_2}{L_2}, \frac{2\pi n_1}{L_1}, \frac{2\pi n_4}{L_4}$ and $\frac{2\pi n_3}{L_3}$ where n_1, n_2, n_3 and n_4 are integers. The discontinuity of $A_1(x)$ ($A_3(x)$) at $x_2 = L_2$ ($x_4 = L_4$) due to the global part only amounts to a gauge transformation. The field tensors F_{12} and F_{34} are continuous on the torus, while other F 's are zero.

To transcribe the continuum gauge fields to the lattice, we take the lattice sites at $x_\mu = 0, a, \dots, (N_\mu - 1)a$, where a is the lattice spacing and $L_\mu = N_\mu a$ is the lattice size. Then the link variables are

$$U_1(x) = \exp [i A_1(x) a] \quad (68)$$

$$U_2(x) = \exp \left[i A_2(x) a + i \delta_{x_2, (N_2-1)a} \frac{2\pi q_1 x_1}{L_1} t^a \right] \quad (69)$$

$$U_3(x) = \exp [i A_3(x) a] \quad (70)$$

$$U_4(x) = \exp \left[i A_4(x) a + i \delta_{x_4, (N_4-1)a} \frac{2\pi q_2 x_3}{L_3} t^a \right] \quad (71)$$

The last term in the exponent of $U_2(x)$ ($U_4(x)$) is included to ensure that the field tensor F_{12} (F_{34}) which is defined by the ordered product of link variables around a plaquette is continuous on the torus. For $U(1)$ gauge field

on a two-dimensional lattice, the link variables are reduced to Eqs. (68) and (69) with

$$A_1(x) = \frac{2\pi h_1}{L_1} - \frac{2\pi Q x_2}{L_1 L_2} + A_1^{(0)} \sin\left(\frac{2\pi n_2}{L_2} x_2\right) \quad (72)$$

$$A_2(x) = \frac{2\pi h_2}{L_2} + A_2^{(0)} \sin\left(\frac{2\pi n_1}{L_1} x_1\right) \quad (73)$$

where the topological charge is

$$Q = \frac{1}{2\pi} \int d^2x F_{12} \quad (74)$$

Acknowledgement

This work was supported by the National Science Council, R.O.C. under the grant number NSC88-2112-M002-016.

References

- [1] S. Adler, Phys. Rev. 177 (1969) 2426; J.S. Bell and R. Jackiw, Nuovo Cimento 60 (1969) 47.
- [2] G. 't Hooft, Phys. Rev. Lett. 37 (1976) 8; Phys. Rev. D17 (1976) 3432; Phys. Rep. 142 (1986) 357.
- [3] For a detailed account and references, see, e.g., S. Weinberg, *The Quantum Theory of Fields*, vol. II, (Cambridge University Press, 1996).
- [4] K. G. Wilson, in *New Phenomena in Subnuclear Physics*, proceedings of the 14th course of the International School of Subnuclear Physics, Erice, 1975, edited by A. Zichichi (Plenum, New York, 1977); K. G. Wilson, Phys. Rev. D 10 (1974) 2445.
- [5] H.B. Nielsen and N. Ninomiya, Nucl. Phys. B185, 20 (1981); B193, 173 (1981).
- [6] T.W. Chiu, Phys. Lett. B445 (1999) 371.
- [7] P. Ginsparg and K. Wilson, Phys. Rev. D 25 (1982) 2649.
- [8] H. Neuberger, Phys. Lett. B417 (1998) 141; Phys. Lett. B427 (1998) 353.
- [9] P. Hasenfratz, Nucl. Phys. B525 (1998) 401.
- [10] S. Chandrasekharan, "Lattice QCD with Ginsparg-Wilson fermion", hep-lat/9805015.
- [11] T.W. Chiu, "Topological phases in Neuberger-Dirac operator", hep-lat/9810002.
- [12] T.W. Chiu and T.H. Hsieh, "Perturbation calculation of the axial anomaly of Ginsparg-Wilson fermion", hep-lat/9901011.
- [13] M. Lüscher, Nucl. Phys. B538 (1999) 515.
- [14] Y. Kikukawa and A. Yamada, Phys. Lett. B 448, 265 (1999).
- [15] K. Fujikawa, Nucl. Phys. B546, 480 (1999).
- [16] D. Adams, "Axial anomaly and topological charge in lattice gauge theory with overlap-Dirac operator", hep-lat/9812003.
- [17] H. Suzuki, "Simple evaluation of chiral Jacobian with overlap-Dirac operator", hep-th/9812019.

- [18] P. Hasenfratz, V. Laliena and F. Niedermayer, Phys. Lett. B427 (1998) 125.
- [19] M. Lüscher, Phys. Lett B 428 (1998) 342.
- [20] C. Callan, R. Dashen and D. Gross, Phys. Lett. 63B, 334 (1976); R. Jackiw and C. Rebbi, Phys. Rev. Lett. 37, 172 (1976).
- [21] R. Narayanan and H. Neuberger, Nucl. Phys. B443, 305 (1995).
- [22] D.B. Kaplan, Phys. Lett. B288, 342 (1992).
- [23] S.A. Frolov and A.A. Slavnov, Phys. Lett. B309, 344 (1993).
- [24] T.W. Chiu, Phys. Rev. D58 (1998) 074511.
- [25] L.H. Karsten and J. Smit, Nucl. Phys. B183, 103 (1981).
- [26] W. Kerler, Phys. Rev. D 23, 2384 (1981).
- [27] E. Seiler and I.O. Stamatescu, Phys. Rev. D25, 2177 (1982); erratum: D26, 534 (1982).
- [28] K. Fujikawa, Z. Phys. C25, 179 (1984).
- [29] I. Sachs and A. Wipf, Helv. Phys. Acta 65, 652 (1992).
- [30] K. Fujikawa, Phys. Rev. Lett. 42, 1195 (1979).

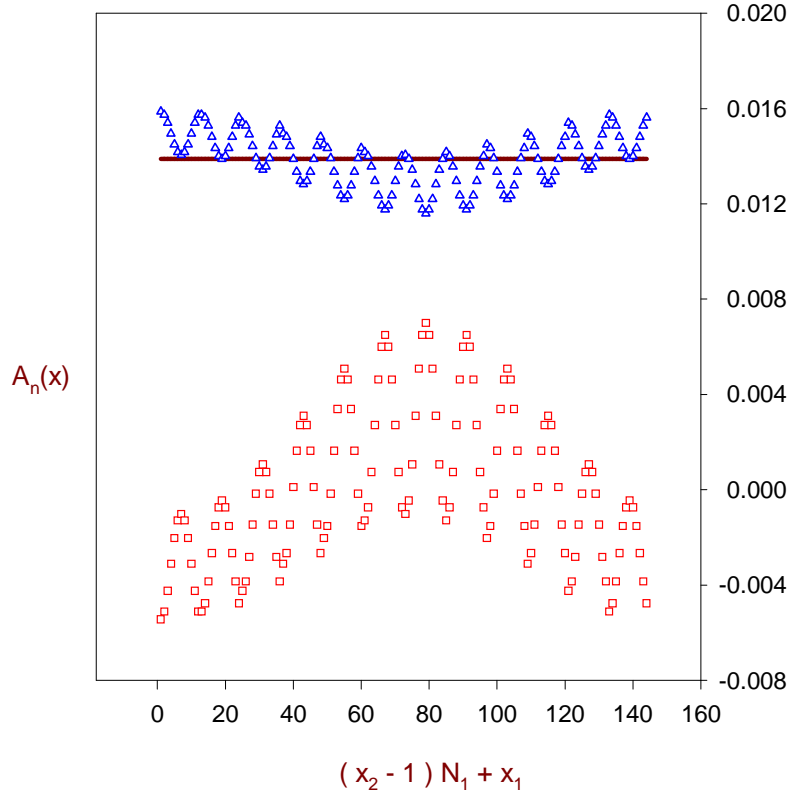


Figure 1: The anomaly function on a 12×12 lattice. The background gauge field has constant field tensor with topological charge $Q = 1$. The horizontal line at height $1/72$ is the Chern-Pontryagin density on the torus. The axial anomaly of D [Eq. (54)] is denoted by triangles which fluctuate around the exact solution, while that of the Wilson-Dirac operator [Eq. (51) with $m_f = 0$ and $r_w = 1$] by squares which are far away from the exact solution. The axial anomaly of D' [Eq. (60)] coincides with the exact solution (the horizontal line at height $1/72$).

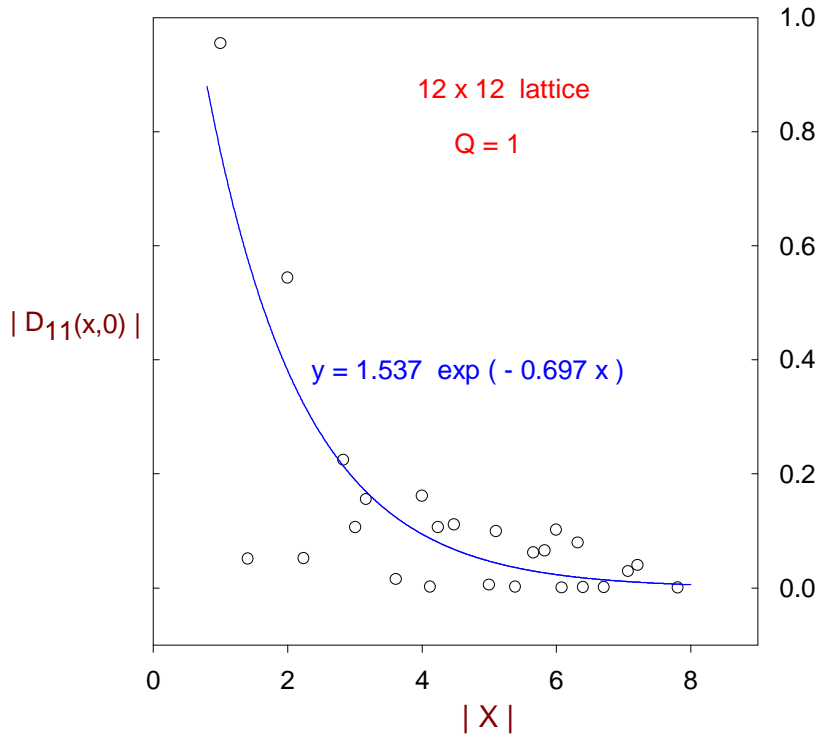


Figure 2: One of the Dirac components of $D(x,0)$, $|D_{11}(x,0)|$, is plotted as a function of $|x|$. The lattice is 12×12 with periodic boundary conditions. The background gauge field has constant field tensor with topological charge $Q = 1$. All data points with the same distance $|x|$ from the origin have been averaged. The solid line is an exponential fit to the data points.

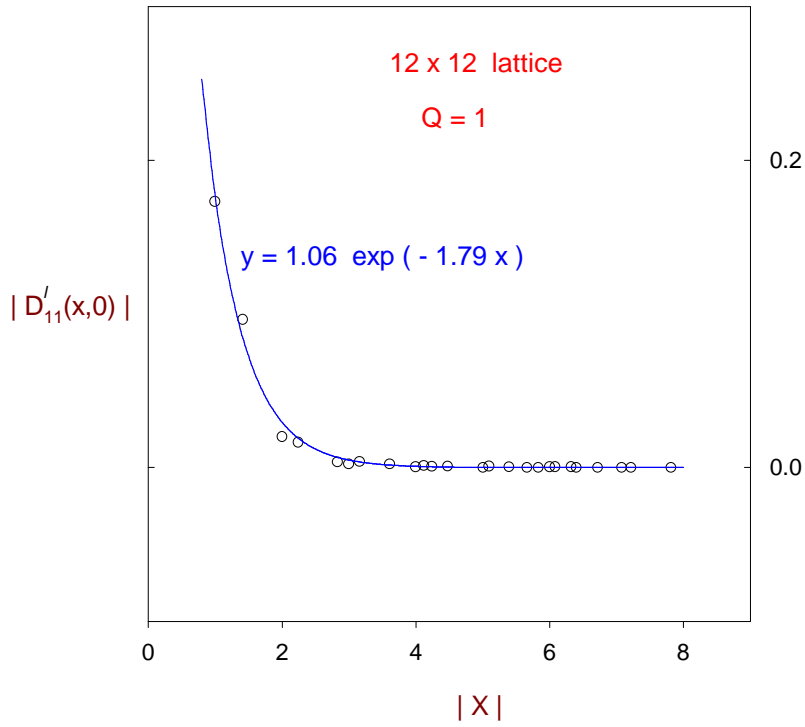


Figure 3: $|D'_{11}(x,0)|$ is plotted as a function of $|x|$. The lattice is 12×12 with periodic boundary conditions. The constant background gauge field has topological charge $Q = 1$. All data points at the same distance $|x|$ from the origin have been averaged. The solid line is an exponential fit to the data points. The same decay constant also fits very well for all other Dirac components of $D(x,y)$ and for any reference point y .

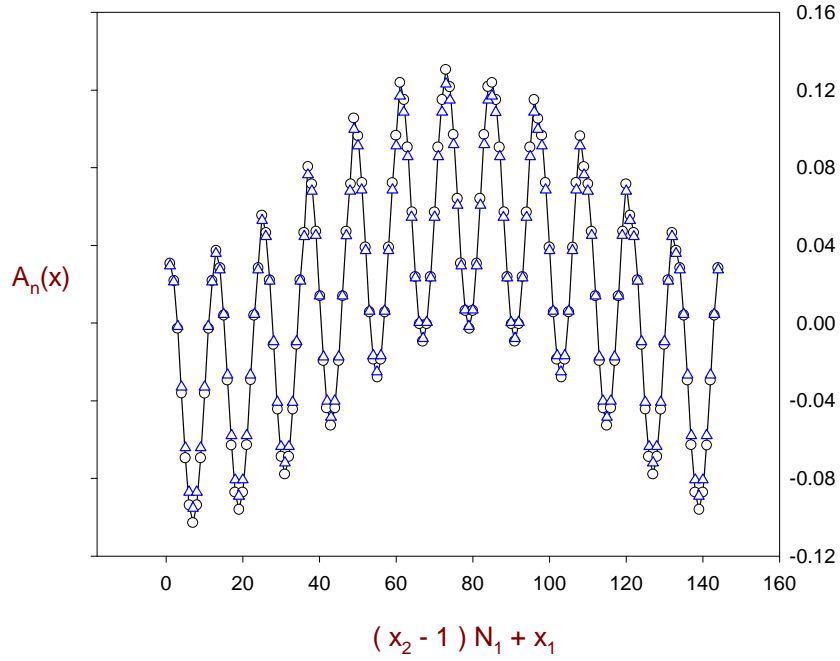


Figure 4: The anomaly function of D' [Eq. (60)] on a 12×12 lattice versus the Chern-Pontryagin density $\mathcal{A}_c(x)$ [Eq. (57)] on the torus. The parameters of the background gauge field are $Q = 1$, $h_1 = 0.1$, $h_2 = 0.2$, $A_1^{(0)} = 0.3$, $A_2^{(0)} = 0.4$ and $n_1 = n_2 = 1$. The Chern-Pontryagin density is denoted by circles which are connected by line segments for the visual purpose. The anomaly function of D' is denoted by triangles.